DOCUMENT RESUME

ED 252 646 UD 024 036

AUTHOR Cooper, Harris M.
TITLE On the Social Psychology of Using Research: The Case of Desegregation and Black Achievement.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE [84]
GRANT NIE-G-82-0022
NOTE 41p.; Some pages may not reproduce well due to light type.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Attitude Change; *Black Achievement; Data Analysis; Data Collection; *Desegregation Effects; *Educational Researchers; Elementary Secondary Education; Evaluation Methods; Graduate Students; Higher Education; Literature Reviews; *Research Methodology; *School Desegregation; Scientific Methodology; Social Science Research
IDENTIFIERS *Research Synthesis

ABSTRACT

The process of research synthesis is examined in this paper. The results of a case study of what happens to the attitudes of six experts in the field of school desegregation and black achievement when they interactively review the relevant research literature are presented. The impact of these subjects' resulting reviews on the attitudes of a less expert audience (i.e. graduate students) is also examined. It is concluded that the attitude-relevant aspects of research reviewing are probably best understood by noting the special characteristics of the people and circumstances involved and by applying established principles to this unique situation. Attitudes are less likely to change if (as in the case of the research reviewers) there is a high degree of prior knowledge. In addition, the diversity of results often found in a set of empirical studies will inhibit attitude change in its consumers, further impeding consensus. And finally, even when there is agreement on the objective outcome of a set of studies, there may still be discrepancies resulting from subjective interpretations of what the measurements mean. Broader philosophic debates within the social sciences which are helping to define the limits of objectivity are considered and further studies based on the premise that knowledge synthesis is more meaningful if accompanied by descriptions of the process that brought the synthesis about are urged. (RDN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

2-5-85

ED252646

On the Social Psychology of Using Research: The Case of Desegregation and Black Achievement

Harris M. Cooper
Center for Research in Social Behavior
University of Missouri-Columbia
111 East Stewart Road
Columbia, MO 65211

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

The research was supported by National Institute of Education grant #NIE-G-82-0022 though the opinions expressed do not necessarily reflect those of NIE. The author wishes to thank Richard Petty, Lee Ross, and Lee Shulman for comments on an earlier draft and Spencer Ward for help throughout the research project.

UD024036

Running Head: Six Desegregation Reviews

Overview

Investigators interested in the social psychology of education often confront a time-worn problem faced by all applied researchers: how can objective and/or consensual rules of data gathering and interpretation be used to study issues that are value-laden and tied in with well-entrenched attitudes? Perhaps this problem is best exemplified by the debate surrounding the effects of school desegregation. The intensity of social and political pressures, the salience of the researcher's personal beliefs, and the complexity of the problem, make desegregation an especially difficult area for interpretation using social psychological principles and techniques. This manuscript presents the results of a case study examining what happens to the attitudes of six experts on the effects of desegregation when they interactively review the relevant research literature. The impact of their resulting reviews on the attitudes of a less expert audience (i.e., graduate students) is also examined.

Introduction

In the summer of 1932, NIE's Desegregation Studies Team (DST) undertook an attempt to reconcile the research literature concerning the effects of desegregation on black children's achievement. The several major reviews in this area had produced conflicting results, leading the Desegregation Studies Team to feel there was a need to clarify what was known about the effects of desegregation, what needs to be discovered, and what directions the next research should take.

Jeffrey Schneider (1982), the head of the Desegregation Studies Team at the time of this effort, has written:

> DST knew from the beginning of the project that much of the available research on school desegregation and its effect on black students' academic achievement suffered from design

> flaws. DST also knew from the beginning that much of this research raised more questions than were answered. We did not know if desegregation research suffers from this in any greater measure than other topical research areas. Nor did we know why so many reputable scholars disagreed in their interpretations of analytical results. To help answer these questions, NIE commissioned a set of papers in an effort to obtain the views of six reputable scholars who had reported opposing conclusions in this area and one research methodologist who had not been identified as a desegregation researcher. The NIE interest in this project was in finding if under similar conditions, with the same set of data, and common ground rules, similarities and differences in scholarly analysis can be identified and clarified. The participants were selected by an NIE staff analysis of previous research findings, by an informal poll of persons engaged in desegregation research, and by a request that each of the possible participants identify others having similar and divergent views from their own. The panel members were Robert Crain, Paul Wortman, David Armor, Norman Miller, Walter Stephan, and Thomas Cook, who served as methodologist.
>
> The seven scholars met in July 1982 at which time they discussed the state of the research literature and agreed on the use of a comprehensive criteria in selecting studies to be analyzed . . . A total of 157 empirical studies were identified that looked at black student academic achievement in desegregated schools. The selection process resulted in a "core" of 19 studies. Panel members agreed, however, that individual reviewers would be allowed to add or delete studies from the "core" (p. 6-8).

After this initial meeting, the panel members prepared first drafts of their papers. They then reconvened in December 1982 to review and evaluate each other's work.

At the same time that the Desegregation Study Team's effort was beginning, NIE's Dissemination in Practice Program was agreeing to fund my research which proposed to study the literature review as a knowledge synthesis process. The first objectives of this research were a naturalistic examination of how literature reviews are carried out and how they are evaluated by interested readers. Obviously, the convening of NIE's panel on desegregation and black achievement provided a rare opportunity for studying the process of research

synthesis. Six expert researchers were asked to draw conclusions about a single hypothesis using a nearly common set of studies. Both the Desegregation Studies Team and the panelists agreed to take part in my research. It is the outcome of this research that will be reported in this paper.

Because of the structure of the panel's assignment, certain aspects of the reviewing process could be examined as part of the naturalistic study. First, each panelist came to the assignment with extensive knowledge of the topic. Nearly all panel members had previously written reviews of desegregation research and had taken part in analysis of primary desegregation data. Therefore, the sources of the experts' predispositions toward the topic could be studied.

Second, the assignment of the panel included a phase in which the quality of desegregation research was to be examined. This allowed for an assessment of the experts' beliefs concerning how a study's design affected its informational utility.

A third set of questions involved attitude change. It was possible to compare the panelists' prior beliefs about desegregation with their beliefs at the assignment's conclusion. An obvious question to ask was: Did the panel experience move the participants toward more positive or negative conclusions concerning the effects of desegregation? Also, did the panel experience enhance or diminish participants' confidence in their conclusions? And perhaps most important, did the panel experience create greater consensus or dispersion among the opinions of participants?

A fourth area of research synthesis that could be examined involved the written products of panel members. The six written

reviews could be read by an interested audience and their reactions to the papers assessed.

The characteristic of the panel that made asking these questions especially meaningful was that all six reviewers addressed a common question yoked to a common body of evidence: All reviewers began with an identical core of 19 studies but were then allowed to add other studies on their own. It is important to note, however, that this same characteristic precluded the asking of some other questions about research synthesis. First, differences in how the reviewers might define the notions of "desegregation" and "achievement" could not be examined. Some panelists found the need for a common definition of these terms restricted their analysis of the problem. Also, variations in literature searching strategies could not be studied because the panel used a pre-existing evidential base. Finally, the panelists all agreed that the generation of an average effect size estimate and the examination of variance in effect sizes across studies was the strategy to be used for integration of study results. Thus, while different reviewers might choose different mathematical formulas, the basic soundness of the quantitative approach was generally accepted.

The factors influencing choices of definitions, literature search strategies and synthesis techniques are as central to the outcomes of reviews as are the factors influencing predispositions, research quality judgments, and report preparation (Cooper, in press). They are overlooked here only because the panel was structured in a manner that meant these parameters of integrative research reviewing were not as free to vary as were others.

Collection of Data

Data collection for the first part of this study was accomplished through two telephone interviews with each panelist. The first interview occurred after the panelists initially met but before actual work on their papers began. This phone interview included closed-ended quantitative scale questions, open-ended questions, and nondirective requests for general observations.

Each participant was asked what their predisposition was concerning the research on desegregation--did it enhance, have no effect, or diminish black achievement, or could no conclusion be drawn? Participants were also asked how confident they were that their interpretation was correct and, in the event that they believed there was a desegregation effect, participants described its magnitude on a scale from "very small" to "very large". The panelists next listed those variables they felt might mediate the effect of desegregation. For instance, among the mediators offered were the child's age at desegregation, curriculum factors, and staff attitudes, to name a few.

In the final part of the structured interview, participants rank ordered six aspects of experimental design with regard to their impact on the "informational utility" of a desegregation study. The six aspects of experimental design included (a) the definition of desegregation employed in the study, (b) the adequacy of the control group, (c) the validity of the achievement measure, (d) the representativeness of the sample, (e) the representativeness of the environmental conditions surrounding the test, and (f) the statistical analysis. So, for instance, if a participant ranked the adequacy of the control group first, it meant he felt this aspect of

the research design had the greatest impact, either positive or negative, on the value of a study's results.

In the second telephone interview, the first three questions were repeated, thus allowing assessments of change in the panelists conclusions, confidence in conclusions, and estimates of effect size magnitude. Participants were also asked about their general political beliefs and, in open-ended questions, about their reactions to the panel experience.

In the second phase of the study, the first drafts of the panelists' written papers were read by 14 post-Masters graduate students in psychology and education. The graduate readers took part in interviews before and after reading the first drafts that paralleled the interviews with the participants. The readers also completed a separate questionnaire concerning each first draft on which they made judgments about the reviewer's positions and the quality of the paper. More details on this phase of the study will be presented after the results of the interviews with the panel members are discussed.

Results: Reviewer Interviews

The first questions of interest involved the sources of the reviewers' predispositions. These impressions were gleaned primarily from the open-ended and nondirective responses of panelists during the phone interviews. I will state some of these results as conclusions general to all reviewers when, in fact, they are really hypotheses for more broadly based and better controlled study.

The most important source of predispositions for panelists was the outcomes of their own primary research. Hands-on experience with

primary desegregation data appeared to form a central set of expectations for the results of any research on the same topic. Seeing--or in this case collecting and analyzing data--is believing. Panelists with primary research experience in the area appeared to give greatest, and perhaps disproportionate, weight to the outcomes of their own studies.

Because the present study was naturalistic and the data on the panelists' initial dispositions was retrospective, the assertion that outcomes of personal primary research caused dispositions is clearly speculative. We must also entertain the notion that initial dispositions led panelists to structure their primary research designs and analyses in a manner that made supportive results highly likely. However, two of the panelists explicitly stated that their beliefs about desegregation effects changed in response to data collection. Probably the most defensible assertion is that both processes exist in nature (that is, primary research influences beliefs and beliefs influence research design and analysis).

Also, there is much evidence that when initial beliefs are vague or tenuous, the impact of data collection will be great. Personal work on primary research is a highly salient event that is considerably more vivid than the research of others. The work of Tversky and Kahneman (1973) indicates that such events are not only overweighted in judgments but often form a preexisting structure (or schema) to which new information is assimilated.

A second source of predisposition was the disciplinary affiliation of the reviewer. Disciplinary affiliations appeard to be most important with regard to the selection of mediators of the

desegration effect. Educators searched mainly for curriculum variables as mediators of the relation, psychologists offered mainly intervening variables associated with interpersonal interaction, and sociologists invoked mainly social structure mediators. Less conclusively, disciplinary affiliations also evidenced themselves in predispositions about results. This may occur because of the reviewer's general faith in his discipline's level of analysis. To use an example other than desegregation, a sociologist might be more likely to agree with the statement "social class affects achievement motivation" than would a psychologist, who might see social class simply as a weak substitute for the "real" psychological determinants of achievement motivation, like the amount of independence training that goes on in the family.

Finally, predispositions appeared to arise from the broader political and social belief systems of the panelists. Many topics in the social sciences relate to real-world problems, and certainly desegregation is one of these. The outcomes of hypothesis tests have implications for the validity of different world views. Conservative and liberal stances on general political issues can "filter" down to imply particular stances on testable research hypotheses. While the present study could only note the consistency, but not causal interrelations, of general belief systems and specific interpretations of empirical data, cognitive consistency theories suggest the pressures toward congruence will work to keep general and specific beliefs consonant with one another (Abelson et al., 1968).

The observation that predispositions toward review outcomes are influenced by personal research, disciplinary affiliation, and general

belief systems is commonsensical. The point needs to be made here, however, because it will shortly become evident that predispositions explained much of the panelists' reactions to their assignment. It appears that a description of the synthesis process would not be complete without clear reference to the role of prior beliefs.

The next set of questions dealt with the impact of research design on the utility of a study's results. Table 1 presents the panelists' rankings. The rankings revealed general agreement that

Place Table 1 about here

design factors associated with internal validity most influence a study's utility. External validity factors were less important. The measurement of the outcome variable, in this case achievement, revealed much variability in its rated importance. It appeared as both the most and least important design factor and only two panelists agreed on its ranking. The statistical analysis used in a study also showed variability in ratings of importance but the ratings were generally low. Several panelists mentioned that if a study had deficiencies in statistical analysis, these could be corrected after the fact.

A Spearman correlation between pairs of rankings revealed an average r of +.47. This number is not dissimilar from correlations found in broader studies of research quality judgments (Gottfredson, 1978). The pair-wise correlations, however, ranged from -.29 to +.77. This finding led to a testing of whether disagreements about the impact of design factors on study utility were associated with disagreements about the effect of desegregation. To do this, participants were

ranked in terms of their perceptions of the effectiveness of desegregation. Then, the difference in ranks between each pair of panelists was correlated with the correlation between their rankings of the design factors. The resulting Spearman r was -.78, indicating that those panelists who disagreed most about the effects of desegregation also disagreed most about relative impact of design factors on study utility. Of course, this is only a crude descriptive device and we cannot determine whether differences in quality criteria account for differences in interpretations of desegregation research, or vice versa, or whether both beliefs are a function of yet a third variable. However, this evidence does support the earlier assertion that dispositions toward specific empirical hypotheses cannot be understood in isolation of broader cognitive structures.

The final set of reviewer data involves the changes in panelists' attitudes toward desegregation. Before the panel began, three participants expressed a belief that desegregation had positive effects and none changed their mind. Two of three panelists who believed desegregation sometimes had no effect and sometimes had positive effects experienced no general attitude change. One participant changed his opinion from this equivocal position to the position that no conclusion could be drawn, due primarily to an enhanced appreciation of the complexity of the issue.

With regard to panelists' confidence in their conclusions, three participants found the experience enhanced confidence in their beliefs. For two participants, this change was dramatic. On a scale from zero (not confident at all) to 10 (totally confident), they moved from 2 to 6 and from 3 to 9. The third panelist showing enhanced

confidence moved from 7 to 9. One participant lost a small, but perceptible, amount of confidence in his conclusion moving from 9 to 7 on the scale. Two participants reported no changes in confidence level remaining at 7 and 8 on the scale.

Next, the panelists' estimated the magnitude of the desegregation effect. Three panelists revised upwa d their estimates of magnitude moving from "very small" to "small", from "small to moderate" to "moderate", and from "moderate" to "moderate to large". One panelist's estimate that the effect was "very small to small" did not change and two participants, who were reluctant to estimate an effect magnitude before the panel began, estimated the effect as between "very small" and "small" when their work was done. These results lead to a conclusion that there was some movement toward more positive impressions of the effect of desegregation. However, it is important to point out that a reading of the panelists' papers, when compared to their earlier writings, generally gives the opposite impression--that is, desegregation seems less impactful in the newer works. This assessment is supported by data on the graduate student readers' impressions before and after reading the papers, to be described shortly.

The conflicting impressions given by the phone interviews and the written papers provide an important insight into the estimation and interpretation of effect sizes. Cooper (1981) has argued that the substantive interpretation of an effect contains two components. The first component is the mathematical estimation of the magnitude of the effect itself. On this estimate, the panelists achieved a high degree of correspondence. All of the panelists agreed that the

effect was positive (i.e., desegregation enhanced black achievement) and the estimates ranged from d = .04 to d = .17, a difference of less than two months of gain on standard achievement tests. The estimate of effect probably captures this mathematical component and indicates that for half the panelists it was larger after the panel experience than before.

The second component of effect size interpretation, however, involves the choice of contrasting elements for purposes of comparison. That is, effects are rarely interpreted in a vacuum. Instead, they are contrasted with effects of a conceptually similar nature. In this case, panelists might ask themselves, "In comparison to other known educational interventions, how effective is desegregation?" It is this interpretation which appears to have been most prevalent in the panelists' written papers and, apparently, several panelists found the desegregation effect did not compare favorably with other forms of educational intervention. Put differently, for some, the panel experience may have raised their expectation concerning what the desegregation effect needed to be in order to be useful. When this second component is added to the mathematical evaluation of effects, the generally positive picture becomes somewhat negative for some panel members.

The estimates of effect also give some clues about whether the panel experience created consensus or dispersion of opinion. Clearly, there was as much disagreement among participants when the panel concluded as when it began, only after the experience, some panelists more firmly held their beliefs. However, there are some mitigating circumstances that offer consolation. First, the issue of

However, there are some mitigating circumstances that offer consolation. First, the issue of desegregation is an exceptionally emotional one involving moral and political stances as well as the scientific perspective. Other issues may prove more amenable to efforts at consensus building through research synthesis and the panel format. Second, the members of the panel brought an unusually large degree of prior experience to their task. Their initial positions were well thought out, complex integrations of knowledge acquired over years of study. Any expectation of dramatic attitude change would have been unrealistic and contrary to research that indicates prior knowledge and experience with an issue makes attitude change more difficult (e.g., Wood, 1982).

Conclusion: Reviewer Interviews

In sum, hours of conversation with the panelists revealed that they probably hold disparate views on many social issues--in fact, they were chosen for participation based partly on their different perspectives. My impression was that the empirical data did create convergence in their thinking on the effects of desegregation though most of this occurred before the panel was convened, when the panelists' attitudes were more malleable. I think the panelists' positions would have been even more diverse had no data or prior synthesis activity taken place.

Finally, a potentially encouraging outcome of the panel concerns something that did not happen. Recent experimental evidence indicates that when people with conflicting beliefs are exposed to a set of studies containing conflicting results, attitudes can become even more polarized (Lord, Ross, & Lepper, 1979). Other research indicates that

further thought about an issue, even in the absence of new information, leads to polarization in the direction of one's initial tendency (Tesser, 1978). There is little evidence that such attitude polarization occurred among the panelists. Two possible explanations for why panelists' beliefs did not show further dispersion can be offered. First, as noted above, the panelists entered this exercise with highly refined and rehearsed positions, limiting the opportunity for change in any direction. The second explanation is less encouraging. While the panelists began with differing opinions on desegregation's effectiveness, none of the panelists held the belief that desegregation had negative effects on black achievement. In other words, panelists disagreed about the existence or magnitude of positive effects while agreeing that the effect was not negative. The Lord et al. (1979) research demonstrated polarization among persons who believed the same intervention had opposite effects. It is impossible to tell whether or not the inclusion of panel members who believed desegregation had negative effects would have revealed evidence for polarization (by, say, the negative group increasing its estimate of the negative effect size or showing enhanced confidence in their beliefs paralleling the changes in confidence demonstrated by the positive panelists).

Results: Reader Reactions

Measures of change. The second phase of this study involved obtaining reader reactions to the six reviews. First, the fourteen graduate student readers were interviewed before reading the reviews. Several questions concerned the readers' educational and topical background and general political beliefs. Four questions were

identical to the repeated measurements obtained from the panelists and were meant to gauge the readers' beliefs about the effectiveness and mediators of desegregation and their confidence in these judgments. Also similar to the panelists', these four questions were readministered after all six reviews had been read. The change in responses to these questions will be examined first.

The graduate readers' general beliefs about desegregation's effectiveness changed little as a function of reading the papers. Eight readers who initially felt desegregation enhanced black achievement did not change their belief, as did one reader who felt desegregation had no effect. Three readers who initially felt no conclusion could be drawn changed to an opinion that the effect was positive while two readers showed change from the latter to the former position. In contrast, the graduate reader's beliefs about desegregation as measured by three other indices changed significantly. Table 2 presents the readers' (a) estimated effect size, (b) confidence in beliefs, and (c) number of suggested mediators of the effect.

Place Table 2 about here

In general, the reader's perception of the positive effect of desegregation dropped precipitously from before to after reading the reviews ($t(12) = -3.97$, $p<.004$). Before reading the reviews, the average effect size estimate was about "moderate", whereas afterward it was less than "small". Eight of 10 readers who estimated the effect size both times revised their estimates downward. Interestingly, the reader's average magnitude of effect was considerably greater than the reviewers' estimate before reading the

papers (4.41 versus 3.00, respectively) but was smaller afterward (2.58 versus 3.16). This reinforces the impression that the papers may have conveyed a less positive evaluation of the desegregation effect than did the reviewer's responses to the interview question.

With regard to the variety of reader opinions, it appears that reading the reviews created some convergence of beliefs--the standard deviation in effect estimates was about two and a half before and one and a third after reading the reviews. However, this finding needs cautious interpretation because two readers who estimated "small" (3) and "moderate" (5) initial effect sizes did not offer an estimate after reading the reviews while two readers who initially left this question blank estimated "very small" (1) effects after reading the papers.

The graduate students' average level of confidence in their beliefs before reading the papers was identical to the reviews' initial confidence levels (6.00). This would indicate that subjective confidence in beliefs about an empirical research area is not directly related (if at all) to objective expertise on the topic. However, after reading the papers, the students' confidence level tended to be higher (7.00; $t(14) = 1.66$, $p<.12$) but the level did not jump as much as did the reviewers' confidence (7.67).

Finally, reading the reviews led graduate students to cite more mediators of the effect of desegregation than they initially proposed, though statistically the effect only approached significance ($t(13) = -2.06$, $p<.07$). While about six and a half mediators were mentioned on average before reading the papers, almost nine were mentioned after reading the papers. Nine readers enhanced the

complexity of their beliefs about desegregation's effect.

It is important not only to note that the readers offered more mediators of desegregation effects as a function of reading the papers, but also to know what kinds of mediators showed the greatest increase in number of citations. In order to find this out, a content analysis was conducted on the open-ended responses of readers to the question "What variables or conditions would you suggest that might mediate the effect of desegregation on black achievement?"

Figure 1 lists the 60 separate categories into which mediators were placed. The 60 categories were then reduced to 11 broader categories for statistical analysis. The 11 categories of mediators

Place Figure 1 about here

were: the attitudes of the people involved in the desegregation effort, including parents, students, teachers, and school administrators; the family and personal background of the students; the geographic and cultural background of the community; the characteristics of the school, of classrooms, and of teachers; the measurement and type of achievement examined in the research; conditions under which desegregation was accomplished; the amount of resources available to a school district; and the number of and response to problems associated with changing students from one school to another. An eleventh, miscellaneous category was used for mediators which coders felt did not fit into any of the previous 10 categories.

The categories were developed by the principal investigator based on a reading of graduate students' responses. Then, each readers' responses were coded by two other judges. Their interjudge reliability

across the 11 categories, as measured by Cohen's Kappa, was .74 when they coded the reader's prior beliefs and .76 when they coded the graduate students' responses to the question after reading the reviews.

Table 3 presents the results of this analysis. First, it should be noted that both prior to reading the reviews and after reading the reviews, the graduate students cited the attitudes of those involved

Place Table 3 about here

in the desegregation effort as the most important influence on its effect. The two types of mediators that evidenced significant increases in citation from before to after reading the reviews were the two categories involving how achievement was defined and how desegregation was accomplished.

Reading the reviews led the graduate students to place an increased emphasis on factors such as how achievement was measured and interpreted, what subject matter was under scrutiny, whether or not the desegregation effort was voluntary or forced, how involved the community was in its implementation, the need or distance of busing, the black-white ratio created by desegregation, the length of time between implementation and the assessment of outcome, and finally whether it was black or white children who got bused. There was also some indication that the readers placed greater emphasis on the need for planning and the avoidance of disruption after the reviews were read.

The next analysis involved relating the readers' background to their beliefs about desegregation before and after reading the reviews. These results are displayed in Table 4. Table 4 contains nine

variables. The variable "topic familiarity" is a composite of the graduate students' year in graduate school and their responses to the

Place Table 4 about here

questions, "How familiar are you with desegregation research?" and "How many scholarly articles related to desegregation have you read?" The intercorrelations among these items were .75, .81, and .83. The three items were only weakly related to the graduate students' answers to the question, "How familiar are you with research methodology?" (correlations ranged from $r = -.10$ to $r = .02$), so this response was left as a separate variable, as was the question asking readers to place themselves on the political spectrum. The last six variables on Table 4 should already be familiar. They are the graduate students' assessments of the direction and magnitude of the desegregation effect, their confidence in their beliefs, and the number of mediators they cited before and after reading the reviews.

In examining Table 4, it should first be noted that the correlations are based on between only 9 and 13 degrees of freedom. Therefore, the tests have very poor statistical power and in some instances I will mention correlations that do not reach traditional levels of significance but are large enough to warrant attention in future studies.

With regard to the relationships between the readers' politics and their desegregation beliefs, no correlation between political beliefs and other variables was found to be significant or to produce a trend. The strongest relationships were between political beliefs and the size of the desegregation effect estimated after reading the

papers. Specifically, the correlation of r = .39 indicates that more liberal readers estimated larger effects after reading the reviews. Before reading the reviews, the relation between political beliefs and effect size was near zero (r = .01). Although both of these figures are nonsignificant, the direction of change in the relation is revealing. If the data were building consensus by replacing abstract beliefs with objective outcomes, we would have expected the opposite pattern of results, that is, a disappearance of a relation between broad political tenets and beliefs about desegregation effects as a function of reading empirical reviews. Instead, something like a polarization effect appears to have occurred in that the readers' political beliefs may have guided their processing and interpretation of the data. Also, in future studies of this issue, it would be wise for researchers to ensure a wider range in initial political beliefs--in the present study readers predominantly described themselves as liberal. A less restricted range of beliefs might reveal a larger initial relationship.

The research expertise of the readers was negatively related to their effect size estimates (r = -.45) and positively related to the number of mediators they cited (r = +.54) after reading the reviews. Thus, it appears that graduate students with more research experience came to see a more equivocal and complex situation as a function of the information input. This is probably what we would expect one of the implications of research expertise to be.

Finally, familiarity with the topic of desegregation was positively related to the graduate students' confidence in their beliefs both before (r = .48) and after (r = .38) reading the reviews,

but the relation was somewhat stronger at the first measurement. Possibly, the mitigated effect is a function of those graduate students who were unfamiliar with the topic showing increased confidence in their beliefs after reading the six reviews. Topic familiarity was also negatively related to effect size estimates before reading the reviews ($r = -.47$). This indicates that those readers least familiar with desegregation research had the higher expectations for desegregation's effect. Interestingly, after reading the reviews, topic familiarity was positively related to effect size estimates ($r = .35$). An examination of the raw data revealed that those readers who were most unfamiliar with desegregation research had their beliefs about effectiveness take the more precipitous drop from before to after reading the reviews. Again, this substantiates the finding that familiarity with a topic is positively related to resistance to change.

Evaluations of individual reviews. The next set of analyses involved the graduate readers' evaluations of the individual reviews. One question asked the reader to estimate the reviewer's position on the direction of the desegregation effect; another question asked what the reviewer would estimate desegregation's effect size to be; and a third question asked the reader to list the reviewer's suggested influences on the impact of desegregation. Readers were also asked to gauge their own confidence in how accurate their interpretation of the reviewer's position was. The second set of questions asked the reader to make seven evaluative judgments about the review. These judgments included how clearly the problem was defined in the review, how exhaustive was the research covered in the review, how well the reviewer evaluated the strengths and weaknesses of desegregation

research, how well the reviewer synthesized the separate studies into a coherent whole, and how clearly the review was written. Readers also gave an overall judgment of the review's quality and persuasiveness. Finally, readers were asked to make open-ended comments concerning their evaluation of the reviews.

The first analysis involved an informal content analysis of the open-ended comments offered by readers at the bottom of each evaluation sheet. This was undertaken to get some idea of the subjective dimensions readers employed to evaluate the reviews, rather than those offered on the questionnaire itself.

The readers most often mentioned that a paper was either well or poorly organized. Second most frequently mentioned was writing style, in particular the author's ability or inability to keep the interest of the reader. Third was how well or poorly focused the paper was on the topic of interest. Fourth was how well or poorly the reviewers used citations to substantiate any claims made in their paper. Next was attention or inattention to variable definitions and to mediating influences. Also mentioned were how well or poorly the reviewer described the methods of the individual desegregation studies and the methods of the review itself. Finally, the manuscript preparation, typically involving negative comments about typos or missing tables, was also mentioned by several of the readers.

To examine the responses to the closed-ended parts of the questionnaires, a factor analysis was first performed on the five judgments concerning specific qualities of the reviews. The factor analyses were performed for each review separately. They revealed that a single quality factor probably underlied all five judgments.

The first principle component in each analysis accounted for between 58% and 83% of the variance in the five scores and other factors had high loadings by only single variables. When the first five questions were standardized and combined into a composite measure of review quality, this composite correlated with the reviewer's overall judgment of quality ranging from .84 to .96 across the six reviews. Therefore, the composite measure of quality based on the first five questions were used in all subsequent analyses rather than the single measure.

The final analysis entailed an examination of the covariation between readers' evaluations of the quality and persuasiveness of a review and their perceptions of the reviewers' beliefs. To carry out this analysis, the readers' responses to the closed-ended items on the individual review questionnaires were correlatd with one another as well as with several of the reviewers' answers to questions during the telephone interviews. These correlations were computed for each reader separately, the correlations for the 14 readers were then transformed to Z-scores and the Z-scores then entered into a one-sample t-test to determine if they were significantly different from zero. An example will clarify this procedure. Each reader made six quality judgments, one for each review, and six judgments of the persuasiveness of the reviews. Therefore, for each reader a correlation was computed, based on six paired observations, which described the direction and strength of the relation between quality and persuasiveness judgments for that reader. This procedure was carried out for each reader and then the set of 14 correlations were transformed to Z-scores to normalize their distribution. Finally, by testing whether the average Z-score was significantly different from zero, it could be established

whether or not a relation between the two judgments could be inferred from the sample of readers.

Figure 2 displays the significant relations between the readers' perceptions of the individual reviews and selected reviewer characteristics. In interpreting these relations, it must be kept

Place Figure 2 about here

in mind that they are associational in nature, so inferences about causal direction and nonspuriousness must be made with extreme caution.

Of foremost interest are those perceptions of reviews that correlated with judgments of quality and persuasiveness. It was found that the quality of reviews was not related to the substantive position of the reviewer, such as the perceived effect size or number of mediators offered. Again, we must bear in mind, however, that the sample of both panelists and readers were all on the same side of the issue (i.e., no one felt desegregation had negative effects). Had a wider range of beliefs been represented, a relation between judged quality and substantive position might have emerged. Instead, reviews judged to be of higher quality were those which the readers felt the most confidence in the accuracy of their interpretation of the reviewers' position, regardless of what that position was.

The persuasiveness of a review was positively correlated with its quality, with the reader's confidence in its interpretation and with the number of mediators the reader thought it mentioned. This last relation, indicating more persuasive reviews were those seen as mentioning more mediators, is of special interest because it indicates a relation between persuasiveness and complexity. It would be

important to know if this relation is general across all topics or if it holds only for topics where readers are predisposed to believe the relation under study is very complex.

The magnitude of the desegregation effect a reader thought a review espoused was positively related to the reviewers' stated effect size and to the liberalness of the reviewers' stated political beliefs. Effect size perceptions were also negatively related to the reviewers' stated confidence in their interpretation and to the number of mediators a reader thought the review mentioned. Finally, the number of mediators a reader thought a review mentioned was positively associated with the reviewers' confidence in their interpretation and with the reviewers' politics.

To summarize the results of the second phase of this study, the graduate readers showed greater flexibility in their attitudes than the reviewers, due undoubtedly to differences in the two groups' initial states of knowledge. In general, readers' beliefs became more congruent with the reviewers' beliefs and more complex as a function of reading the papers. What was not examined, however, is whether under natural circumstances readers would have chosen to expose themselves to the variety of opinions contained in these papers or whether they would only seek out papers that would confirm or bolster their initial positions. This latter phenomenon, called selective exposure, is a source of controversy in social psychological research.

The analysis of readers' backgrounds found little initial bias due to political beliefs but some indication that increased knowledge of the topic also led to increased congruence between general political beliefs and beliefs about desegregation. This may indicate the

presence of the phenomenon mentioned earlier--that when people are confronted with a smorgasbord of empirical findings they give greatest weight to those that are consistent with prior beliefs.

With regard to the evaluation of individual reviews, these appeared to be relatively free from the influence of initial topical beliefs. Quality judgments covaried with confidence in interpretation. Judgments of the persuasiveness of a review covaried with quality, confidence in interpretation, and number of mediators mentioned.

Conclusions

The basic principles of attitude change appear to apply to the process of empirical knowledge synthesis. Rather than being an activity of a qualitatively different order, the attitude-relevant aspects of research reviewing, as exemplified by NIE's panel, are probably best understood by noting the special characteristics of the people and circumstances involved and by applying established principles to this unique situation.

First, the desegregation panelists began their synthesis task with a great deal of prior knowledge, as do most research reviewers. This knowledge is not only great in magnitude but it is also well organized in a complex structure. These characteristics of reviewers will lessen the possibility of change in basic beliefs because the synthesizer will (a) encounter few arguments which are truly novel and (b) have a cognitive schema with which to integrate or counterargue information that is new.

Readers of research reviews who bring less tenacious beliefs to the topic area, such as the graduate students in this study, are more

likely to experience attitude change and enhanced complexity of beliefs. The amount and direction of this change apparently will not only be a function of the reviewer's conclusions and treatment of the relevant material, but also of the effectiveness of the reviewer's presentation, in particular the organization and style of the manuscript.

Aside from the characteristics of the actors in the review process, the empirical character of research syntheses, like the question of the effects of desegregation, is also critical to understanding related attitude change. Scientific studies claim a certain degree of objectivity for their results, and by implication assume that identical tests of the same hypothesis will lead to identical, or at least similar, results. Yet, sets of empirical studies yield varying, often conflicting findings. This may be because methods are different, hypotheses really are not commensurate, or the assumption of objectivity is false at the start.

The diversity of results often found in a set of empirical studies will inhibit attitude change in its consumers. If an initial opinion is minimally reasonable, an examiner of the related research will find some studies that confirm the initial belief. Obviously, the diversity in results will also greatly impede the ability of research to create consensus among reviewers or review readers. In the instance of less knowledgeable consumers, like the graduate readers in this study, the diversity of results may lead to the formation of attitudes consistent with other more generally held beliefs (e.g., political philosophy).

Finally, even when a certain degree of consensus is reached on

the objective outcome of a set of studies, in this case the size of the desegregation effect, the varying perspectives of reviewers and readers can still create discrepancies in the subjective utilities that are used to interpret the findings. Thus, while a great deal of agreement might be reached on the observation that an eight-ounce glass contains four ounces of water, there can still be much disagreement about whether the glass is half empty or half full.

It is this last point which speaks most directly to studying the social psychology of education in general. Many problems in education require both objective measurements and subjective interpretations of what the measurements mean. Questions like "Is a teaching method that fosters independent learning better than one that maximizes achievement?" or "Is the cost of a program justified by its academic outcomes?" are not answerable with data alone. Researchers cannot resolve these debates--they can only help the debaters ground their arguments in more precise and reliable information (e.g., How much independence is gained by a teaching method? How much added knowledge does a program buy?). When the researcher moves from informing debaters to joining the debate, they move from the role of researcher to that of advocate.

Of course, the distinction between these roles is not as clear as it sounds. Often the choice of problems and how they are to be studied can be as value-laden as how data are interpreted. For instance, in assessing the effects of desegregation, its impact on white as well as black student's achievement, or on housing patterns might lead to different conclusions about the treatment's effectiveness. Thus, social psychologists of education will be

challenged at every stage of a research endeavor to remain aware of when they are acting as knowledge-seekers and when as advocates.

If this assessment of the research synthesis process in educational research seems disheartening it is only because the expectations for modern-day social science may be too high. It is unreasonable to consider a failure anything less than complete agreement among scholars. It is also unreasonable to overlook those aspects of a problem upon which agreement has been reached: NIE's panelists achieved remarkable consensus on the magnitude of the desegregation effect.

Also, it should be remembered that the development of new techniques for increasing the objectivity of research synthesis, such as computerized literature searching and meta-analysis, is occupying the time of many social science methodologists. Using these techniques, as the panelists did, probably led them to greater agreement than would have been the case otherwise. It is likely that further refinements in method will lead to more consensus and, just as important, better understanding of why disagreement occurs.

Finally, philosophic debates within the social sciences are helping to define the limits of objectivity, and as such will help us know when a research synthesizer is talking as a scientist and when as a citizen.

In sum, the convening of NIE's panel on desegregation and black achievement proved to be an exciting natural laboratory for the study of knowledge synthesis. It would serve the purposes of both the social science community and the general public if efforts of this sort were continued in the future. Social scientists would learn about basic

social psychological processes and how they influence their own activities. The general public would learn more about what to expect from social science products and how best to interpret them. Knowledge synthesis is more meaningful if it is accompanied by descriptions of the process that brought the synthesis about.

References

Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., and Tannenbaum, P. H. (Eds.), Theories of cognitive consistency: A source book. Chicago: Rand McNally, 1968.

Cooper, H. M. On the significance of effects and the effects of significance. Journal of Personality and Social Psychology, 1981, 41, 1013-1018.

Cooper, H. M. Integrative research reviewing: A social science approach. Beverly Hills: Sage, in press.

Gottfredson, S. Evaluating psychological research reports. American Psychologist, 1978, 33, 920-934.

Lord, C., Ross, L., and Lepper, M. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of Personality and Social Psychology, 1979, 37, 2098-2109.

Schneider, J. Overview of research on effects of desegregation on achievement. In P. Wortman (Chair), An analysis of methodologies used in synthesis of research on desegregation and student achievement. Symposium presented at the meeting of tne American Educational Research Association, Montreal, 1983.

Tesser, A. Self-generated attitude change. Attitudes in experimental social psychology. New York: Academic Press, 1978.

Tversky, A., and Kahneman, D. Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 1973, 5, 207-232.

Wood, W. Retrieval of attitude-relevant information from memory: Effects on susceptibility to persuasion and on intrinsic motivation. Journal of Personality and Social Psychology, 1982, 42, 798-810.

Table 1

Impact of Research Design Factors on the Utility of a Study's Results

	*	*	*	*	*	*	$\bar{x}$
Experimental Manipulation (Definition of Desegregation)	1	1	3	2	1	2	1.6
Experimental Comparison (Adequacy of Control Group)	3	2	2	1	2	1	1.8
Outcome Measure (Measurement of Achievement)	2	6	1	3	4	6	3.6
Population Generality	5	5	4	4	5	5	4.6
Ecological Generality	5	3	5	6	3	4	4.3
Statistical Analysis	4	4	6	5	6	2	4.5

Notes: $\bar{r}$ = +.47, ranging from r = -.29 to +.77

Table 2

Graduate Reader Beliefs About Desegregation Before and After Reading the Reviews

Question:	Before	After
What is the size of the desegregation effect?	4.41 (2.57)	2.58 (1.38)
How confident are you that your belief is accurate?	6.00 (1.92)	7.00 (1.36)
What variables of conditions mediate the effect? (number mentioned)	6.46 (3.45)	8.93 (5.70)

Note: Standard deviations are in parentheses.

Table 3

Change in the Number of Mediators Mentioned by Readers in Eleven Categories

Mediator Category	Prior	Post	t-value	p-level
Attitudes	1.71	1.86		
Student Background	1.11	1.39		
Community	0.86	0.56		
School Characteristics	0.32	0.38		
Classroom Characteristics	0.11	0.25		
Teacher Characteristics	0.46	0.86		
Achievement Definition	0.14	0.79	3.23	.007
Desegregation Definition	0.96	1.75	2.27	.04
Money	0.46	0.25		
Change Problems	0.07	0.25	2.11	.06
Other	0.14	0.57		

Table 4

Relations Between Reader's Background, Prior- and Post-Beliefs About Desegregation[a]

	Topic Familiarity	Research Expertise	Political Beliefs[b]	Prior Effect Size[c]	Prior Confidence	Prior Mediators	Post Effect Size[c]	Post Confidence	Post Mediators
Topic Familiarity	---	-.04	-.02	-.47*	.48*	-.30	.35	.38*	-.24
Research Expertise		---	.05	-.01	-.06	.10	-.45*	.00	.54**
Political Beliefs			---	.01	.08	-.36	.39	-.35	.15
Prior Effect Size				---	.12	.51*	.01	-.53*	.07
Prior Confidence					---	.06	.48*	.09	-.38
Prior Mediators						---	-.12	-.19	.58**
Post Effect Size							---	-.22	-.38
Post Confidence								---	-.26
Post Mediators									---

Notes: * .15<p>.05
** p<.05
a. All tests are based on between 9 and 13 df.
b. More liberal beliefs were given higher numerical values.
c. Effect sizes were coded as negative values if the reader thought desegregation had negative effects. If absolute values were used, one trend emerged - r with prior confidence = .45*.

BEST COPY AVAILABLE

Figure 1

Coding Frame for Moderators of Desegregation Effects

I. Attitudes (support)
of parents, white, black; of students, white, black; of teachers; of school administration; of politicians; of business; of clergy.

II. Student Background
parent education; family relations; family size; social class, blacks, whites; achievement/ability, blacks, whites; age; self concept/personality.

III. Community Background
area of country; urban vs. rural; size; historical conditions; media presentation.

IV. School Characteristics
size; location; quality (curriculum).

V. Classroom Characteristics
size (teacher/pupil ratio; # of teachers); seating (interaction) patterns; open vs. traditional.

VI. Teacher Characteristics
ability; labelling (expectation) effects; training for desegregation; race.

VII. Achievement Definition
measurement; subject matter; interpretation of measurement.

VIII. Desegregation Definition
voluntary vs. forced; community involvement in implementation; need (distance) for busing; black/white ratio; length of implementation; who gets bused.

IX. Money (expenditures)
resources for teachers; other support services (personnel).

X. Change of School Problems (confusion)
preparation of students.

XI. Other

Figure 2

Significant Relations Between Reader Perceptions of Individual Reviews and Reviewer Characteristics

The quality of a review was correlated

(a) positively with the reader's confidence in its interpretation ($\bar{Z}$ = .42, p<.02)

The persuasiveness of a review was correlated

(a) positively with the reader's judgment of its quality ($\bar{Z}$ =1.05, p<.0001)
(b) positively with the reader's confidence in its interpretation ($\bar{Z}$ = .66, p<.001)
(c) positively with the reader's judgment of the number of mediators it mentioned ($\bar{Z}$ = .32, p<.05)

The effect size in a review was correlated

(a) positively with the reviewer's stated effect size ($\bar{Z}$ = .92, p<.0001)
(b) positively with the reviewer's politics ($\bar{Z}$ = .32, p<.02)
(c) negatively with the reviewer's own confidence in their interpretation ($\bar{Z}$ = -.76, p<.04)
(d) negatively with the reader's judgment of the number of mediators it mentioned ($\bar{Z}$ = -.92, p<.01)

The number of mediators in a review was correlated

(a) positively with the reviewer's confidence in their interpretation ($\bar{Z}$ = .62, p<.0001)
(b) positively with the reviewer's politics ($\bar{Z}$ = .55, p<.0001)

Note: All data are correlational but are repeated under only one heading for ease of presentation.